AD_____

Award Number: DAMD17-99-1-9267

TITLE: Long Interspersed Nuclear Element (LINE) Mediated Genomic Instability

PRINCIPAL INVESTIGATOR: Lalitha Nagarajan, Ph.D.

CONTRACTING ORGANIZATION: The University of Texas
M. D. Anderson Cancer Center
Houston, Texas 77030

REPORT DATE: July 2002

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2002	Final (1 Jul 99 - 30 Jun 02)

4. TITLE AND SUBTITLE
Long Interspersed Nuclear Element (LINE) Mediated Genomic Instability

5. FUNDING NUMBERS
DAMD17-99-1-9267

6. AUTHOR(S) :
Lalitha Nagarajan, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of Texas
M. D. Anderson Cancer Center
Houston, Texas 77030
E-MAIL: lnagarajan@notes.mdacc.tmc.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
Original contains color plates. All DTIC reproductions will be in black and white.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

none provided

14. SUBJECT TERMS
breast cancer, genomic instability, chromosomal gain, 1q41, homeobox protein, novel tumor suppressor gene family and chromosomal loss 1p32, 5q13 and 19p13.2

15. NUMBER OF PAGES
29

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

N/A Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

N/A Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

N/A For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Lalitha Nagarajan 11/7/02

PI - Signature Date

Table of Contents

INTRODUCTION

The overall goals of this proposal are to isolate target oncogenes, tumor suppressor genes and regions of chromosomal instability by the use of a long interspersed nuclear element based PCR assay. We fully exploited the near completion of the human genome sequencing with bioinformatic tools to achieve our goals.

We targeted both known and novel genomic loci for potential candidate genes. The first region was the c-Met oncogene, because of its association with poor prognosis in node negative breast cancer (1). The second line of investigations on a novel sequence specific single stranded DNA binding protein (SSBPs) stemmed from our parallel studies on human leukemia with poor prognosis. Originally isolated in the chicken as a gene whose product binds pyrimidine rich mirror repeats elements the existence of three members that map regions of chromosomal instability was unknown until our studies uncovered this novel family. Our investigations also centered on characterization of the protein products for the human *MIXL* gene and the SSBP2 gene product in breast cancer cell lines. Our progress with these promising candidate genes is likely to yield biologically and clinically relevant reagents to diagnose breast cancer with enhanced sensitivity.

Body

For additional technical details please refer to the two publications (Castro et. al.,2002, Guo et. al., 2002) generated through this DOD award in the appendix.

I. Isolation of tumor-specific targets of L1Hs retrotransposition and recombination sites in mammary carcinogenesis.

In accomplishing the tasks i-iii of Specific Aim I, we have taken advantage of the rapid progress made in the human genome sequencing initiative and the cancer genome anatomy project. Thus we have screened the expressed sequence tagged database (dBEST) electronically for transcripts with repetitive sequences.
About the same time, other studies in our laboratory were searching for the human homolog of the homeobox gene *Mix.1* which regulates mesoderm and endoderm differentiation in Xenopus embryos. A single candidate EST clone (AA84781) with a putative *Mix* like homeodomain was identified in a library from germinal center B lymphoid cells. When the EST clone AA84781 was sequenced, we identified a partial open reading frame. A recursive search of the dBEST database identified only two other cDNA clones for this novel gene, suggesting an extremely restricted pattern of expression. Interestingly, one of these ESTs (AA911377) was from a cDNA library made from laser captured infiltrating ductal carcinoma of the breast (Guo et. al., 2002). More importantly the 3' (> 1.0kbp) UTR of this gene contains alu and MER like repetitive elements.

(iv)Sequence analysis and development of STSs

We developed unique STSs from the 3'untranslated region of the MIX like homeodomain gene. Using these primers we screened a human genomic library in Bacterial Artificial Chromosome vector to isolate a large segment of the genome.
The BAC was used to characterize the full length open reading frame of this novel homeogene that we designated *MIXL2* . The gene encodes an open reading frame of 233 amino acids and shows 77% homology to a similar gene, *mml* in the mouse.

Additionally, we characterized the genomic organization and expression pattern of MIXL2. The gene consists of two exons and shows a highly restricted expression pattern. None of the breast cancer cell lines tested showed abundant transcripts. However, reverse transcription coupled PCR detected MIXL2 transcripts in the cell line MDA-MB 453 and not in the immortalized breast cell line MCF10A.

Simultaneously, we also localized the novel gene to human genome using DNA from a panel of well characaterized radiation hybrids. Interestingly, the MIX like gene localized to human chromosome 1q32-41 locus, between the microsatellite loci D1S479 and, a region of gain in breast cancer.

(v) Development of the PCR for minimal material

In collaboration with Dr. Nour Sneige in the Dept. of Pathology, we dissected 25 breast tumors and isolated normal and tumor DNAs. The status of D1S479 was examined in these paired samples by polymerase chain reaction (PCR) with end labeled primers. Interestingly, 12 out of 25 tumors showed allelic imbalance for D1S479 suggesting that the MIXL gene was a candidate target gene of chromosomal gain in 1q

(vi) Correlation of data from tasks (i-v) with stage and grade.

All of the 25 breast tumors used to screen for allelic imbalance of *D1S479* showed lymph node involvement. Studies are underway to determine whether there is intra-tumoral heterogeneity for gain of 1q32-41 loci. We will utilize both FISH and MIXL specific antibody based immuno fluorescence staining approaches to examine such a possibility.

Tasks v and vi of Specific Aim 1 were to Develop PCR for minimal material and correlate data from tasks (i-v) of specific aim 1 with stage and grade to determine consistent quantitative or qualitative patterns can be discerned. Another part was to complete specific Aim 2 which was to localize PCR rescued fragments to chromosomal regions and then identify whether these loci are targets of allelic imbalance in breast cancer.

With the near completion of the human genome sequencing studies under this specific aim were re-focused and approached in a cost and time effective manner by utilizing bioinformatic tools. In addition, we also expanded the scope of the proposed studies by initiating experiments on the protein products of potentially important candidate genes. Thus the current report comprises of two parts: (1) Chromosomal localization of regions of instability and identification of candidate genes and (2) Development of antibodies and characterization of expression of potential tumor promoting and tumor suppressor genes in breast cancer cell lines.

II.Characterization of the genomic locus spanning the tumor specific rearrangements.

A search of the human genome database for novel LINE rich segments showed a high LINE content on chromosome 7q31.1 The *Met* oncogene encoding a tyrosine kinase receptor, and a marker for poor prognosis in breast cancer localizes to this region. Therefore, we searched the genomic sequences in the public domain for putative polymorphic sequences. A Bacterial Artificial Chromosome 354 L07 had a

> 60% repetitive sequences and an unusually high number of L1Hs elements. We developed three novel markers from these sequences designated ca1, ca2 and gata

In collaboration with Dr. Nour Sneige in the Dept. of Pathology, we dissected 25 breast tumors and isolated normal and tumor DNAs. We developed PCR conditions to detect allelic imbalance for ca1, ca2, gata in minimal patient material. These markers along with 7 other highly polymorphic markers (designated D7S***) were used in allelo typing studies of 25 breast tumor and matched control DNAs. 13 out of 25 tumors examined showed allelic imbalance. We could identify two critical regions, one flanked by the C- MET protooncogene and second a more centromeric region between the loci *D7S471* and *D7S1817*.

All of the 25 breast tumors used to screen for allelic imbalance of 7q31.1 showed lymph node involvement. We will interrogate the clinical correlation of both the previously identified chromosome 1q41-42 locus as well as additional loci characterized here once we have conditions optimized to detect the candidate gene expression by immuno staining.

III. Molecular cloning of target genes/loci from the regions of instability.

A novel family of sequence specific single stranded DNA binding proteins localize to regions of allelic imbalance:

Another novel family of evolutionarily conserved genes which localize to regions of genomic instability were identified in our laboratory as part of our studies on human leukemia. The three members of this family encode a putative sequence specific single stranded DNA binding activity (SSBPs2-4). More importantly these map to regions of deletions in breast cancer (Castro et.al., 2002). SSBP3 maps to chromosome 1p32, SSBP2 to chromosome 5q13.3 and SSBP4 to chromosome 19p13.1.

In addition, we also examined the genomic organization of SSBP2 which localizes near a chromosomal break point in the breast cancer cell line SKBR3, as well as a human leukemia cell line ML3. The gene is encoded by 17 exons and there are two large introns (> 100kbp), 1 and 4 with several copies of L1 Hs elements. Interestingly SSBP3 and 4 also have identical genomic organizations with 17 exons although the intron sizes are smaller.

Development of antibodies, characterization in breast cancer cell line models

A corollary of our hypothesis is that altered expression, either enhanced expression in the case of oncogenes or loss of expression in the case of putative tumor suppressor genes at the site of chromosomal instability, contribute towards the phenotypic evolution of breast cancer cells, we developed highly specific

antibodies against peptide epitopes of both MIXL and SSBP2 gene products. Two different epitopes were identified and affinity purified antisera were generated. The specificity of the antibodies were verified in cell lines induced to over express under transient transfection conditions. Thus the antibodies were evaluated stringently and then utilized in immunoblotting experiments with human breast cancer cell lines.

MIXL2 protein is expressed in immortalized and malignant mammary cell lines:

As stated earlier expression of MIXL appears to be highly restricted by RTPCR and Northern blotting analyses. This experimental evidence is confirmed by Bioinformatic analyses of cDNA sequences in the public domain. To date the dBEST Database of more than 10000 entries from over 30 tissues, contains only 4 sequences for the MIXL2 gene. These are from a infiltrating ductal carcinoma in situ, germ cell tumor, normal pre B lymphoid cells and the highly metastatic fibrosarcoma cell line HT1080. Interestingly, all the breast cancer cell lines (SKBR3, MDA-435, MDA- 453,MDA-468 and MCF 7) and the SV40 immortalized mammary epithelial cell line MCF10A express the MIXL protein. This is in contrast to other malignancies including hematopoietic cancer where we detect a highly restricted expression. Thus aberrant MIXL expression might confer a proliferative/ survival advantage to breast cancer cells.

SSBP2 protein is not expressed in immortalized and malignant mammary cell lines:

The putative sequence specific single stranded DNA binding proteins in themselves may be involved in genomic stability as they are postulated to bind single stranded sequences that loop out when mirror repeat elements in the DNA assume a triple helical confirmation. Loss of expression of these elements that reside at regions of genomic instability and eletion, may confer a survival/ loss of ability to repair DNA damage advantage to malignant cells. None of the breast cancer cell lines as well as the immortalized breast epithelial express the SSBP2 protein. This is in contrast to certainT lymphoid cell lines which express abundant SSBP2 protein. Our future studies will elucidate the mechanism by which loss of SSBP2 expression confers survival advantage in cancer.

Key Research Accomplishments:

- Identification of a novel regulatory gene with repetitive elements in the 3'UTR
- Isolation and characterization of the full length open reading frame for the novel gene designated *MIXL2*
- Localization of *MIXL2* to 1q41, a region of gain in breast cancer

- Detection of allelic imbalance for *D1S479*, a marker physically linked to *MIXL2*, in 12/25 breast tumors with lymph node involvement
- Characterization of the 7q31.1 loci flanking the *MET* protooncogene
- Identification of a novel family of sequence specific single stranded DNA binding protein (SSBPs2-4) and localization to human genome
- Development a PCR assay to detect expression of SSBPs2-4
- Development of MIXL specific antibodies
- Detection of MIXL expression in breast cancer cells in culture
- Development of antibodies to SSBP2
- Lack of SSBP2 expression in breast cancer
- SSBP2 binds pyrimidine rich mirror repeat elements in single stranded DNA.
- SSBP2 inhibits HrasV12 mediated transformation of NIH 3T# fibroblasts.

(8) Reportable outcomes:

Development of FISH and immunohistochemical staining probes:

We have characterized a BAC probes from chromosomes 1p32.1, 5q13.3, 7q31.1 and 19p13.1 that can be used in Fluorescence in situ hybridization.

We have developed highly specific antibodies to MIXL2 and SSBP2, which can be used in immuno histochemical staining.

Informatics:

The full length MIXL 2 sequence has been submitted to Genbank.

Full length cDNA and exon specific genomic sequences for SSBPs 2-4 are available in Genbank

Funding applied for based on work supported by this award:

1. Some of the results obtained in this award contributed to the development of an amended RO1 proposal to NIH entitled "SSBP2 gene pathway in Myeloid Neoplasm", submitted Nov.1, 02.

2.Chromosome 5q13 Tumor Suppressor Locus In Prostate Cancer- Candidate Gene *SSBP*2(PI Nagarajan)
Agency: DOD Prostate Cancer Research Program
Duration: 1/01/03 - 12/31/05.
Level of Funding requested:$ 367,186

3.MIXL: A NOVEL HOMEOBOX GENE ACTIVATED IN BREAST CANCER
Time Commitment:10%
Agency: DOD BCRP Idea award
Duration: 1/01/03 - 6/30/04.
Level of Funding requested:$ 297,852

4.*SSBP* a novel tumor suppressor gene family inactivated in breast cancer
Time Commitment:10%
Agency: DOD BCRP Exploration award
Duration: 1/01/03 - 6/30/04.
Level of Funding requested:$ 146,676

Conclusions:

1) In breast caner, there are two critical regions of allelic imbalance centromeric of the c-Met oncogene. Whether imbalances of these intervals are selected for or result from an overall genomic instability remains to be investigated.

2) The sequence specific single stranded DNA binding proteins are encoded by a novel gene family. All three members localize to regions of chromosomal deletions in breast cancer. The SSBP2 gene product inhibits H-ras V12 mediated transformation of NIH 3T3 fibroblasts.

3) A novel paired type- homeodomain protein Mixl2 is expressed in immortalized as well as transformed mammary epithelial cell lines.

REFERENCES

Castro P, Liang H, Liang JC, Nagarajan L. A Novel, Evolutionarily Conserved Gene Family with Putative Sequence-Specific Single-Stranded DNA-Binding Activity Genomics 80:78-85, 2002.

Guo W., Chan AP., Liang H., Wieder ED, Molldrem JJ, Etkin LD and Nagarajan L. A Human Mix-like Homeobox Gene *MIXL* Shows Functional Similarity to *Xenopus* Mix.1 Blood 100:89-95, 2002.

Article doi:10.1006/geno.2002.6805, available online at http://www.idealibrary.com on IDEAL

A Novel, Evolutionarily Conserved Gene Family with Putative Sequence-Specific Single-Stranded DNA-Binding Activity

Patricia Castro,[1] Hong Liang,[1] Jan C. Liang,[2] and Lalitha Nagarajan[1,*]

[1]*Department of Molecular Genetics, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA*
[2]*Division of Pathology and Laboratory Medicine, The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, USA*

**To whom correspondence and reprint requests should be addressed. Fax: (713) 792-0981. E-mail: Lalitha@odin.mdacc.tmc.edu.*

Complete and partial deletions of chromosome 5q are recurrent cytogenetic anomalies associated with aggressive myeloid malignancies. Earlier, we identified an ~ 1.5-Mb region of loss at 5q13.3 between the loci *D5S672* and *D5S620* in primary leukemic blasts. A leukemic cell line, ML3, is diploid for all of chromosome 5, except for an inversion-coupled translocation within the *D5S672–D5S620* interval. Here, we report the development of a bacterial artificial chromosome (BAC) contig to define the breakpoint and the identification of a novel gene *SSBP2*, the target of disruption in ML3 cells. A preliminary evaluation of *SSBP2* as a tumor suppressor gene in primary leukemic blasts and cell lines suggests that the remaining allele does not undergo intragenic mutations. *SSBP2* is one of three members of a closely related, evolutionarily conserved, and ubiquitously expressed gene family. *SSBP3* is the human ortholog of a chicken gene, *CSDP*, that encodes a sequence-specific single-stranded DNA-binding protein. *SSBP3* localizes to chromosome 1p31.3, and the third member, *SSBP4*, maps to chromosome 19p13.1. Chromosomal localization and the putative single-stranded DNA-binding activity suggest that all three members of this family are capable of potential tumor suppressor activity by gene dosage or other epigenetic mechanisms.

Key Words: chromosome 5q13.3, myelogenous leukemia, loss, growth, differentiation, suppression

INTRODUCTION

On the basis of the paradigm that chromosomal deletions represent loss of regions harboring tumor suppressor elements, we and others have searched for candidate suppressor genes on human chromosome 5q31, a common target of deletion in myeloid malignancies. Most of the deletions appear to overlap at 5q31.1; nonetheless, these large interstitial deletions encompass > 60% of the long arm. Moreover, high-resolution cytogenetics has identified a subset of patients who lose only the proximal 5(q11q13) interval and retain the rest of chromosome 5q [1]. Consistent with the cytogenetic findings, we identified a second critical locus at 5q13.3 that is lost from unbalanced translocations between chromosome 5q and 17p. The translocations in two patients with acute myelogenous leukemia (AML), which were in opposite orientations, overlapped within an interval flanked by the markers *D5S672* and *D5S620* [2–4].

The identification of an AML cell line model underscored the importance of the loss or disruption of the 5q13 interval. Specifically, in the cell line ML3, the entire chromosome 5 sequences are grossly intact, except for an inversion-coupled translocation at 5q13.3 within an estimated distance of 1 Mb between the loci *D5S1464* and *D5S620* [2,4]. The *D5S1464–D5S620* interval is contained within the telomeric half of the *D5S672–D5S620* critical locus originally identified in patients. Here we report our findings of a novel gene, *SSBP2*, that encodes a putative sequence-specific single-stranded DNA-binding protein. *SSBP2* is the apparent target of unbalanced translocations and deletions. Additionally, two other members of this family, *SSBP3* and *SSBP4*, localize to chromosome 1p31.3 and 19p13.1, which are regions of deletions in a variety of malignancies.

RESULTS

Physical Map of the 5q13.3 Locus and Delineation of the Chromosomal Breakpoint in the AML Cell Line ML3

Both conventional cytogenetics and fluorescence *in situ* hybridization (FISH) studies of ML3 cells have identified a

0888-7543/02 $35.00

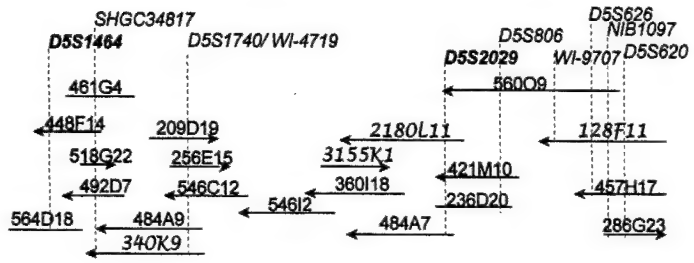

FIG. 1. Minimal tiling path between markers *D5S1464* and *D5S620*. A bidirectional walk between *D5S1464* and *D5S620* was carried out by PCR screening Caltech-C (CTC) BAC library pools. Ends of individual BACs were sequenced, and overlapping clones were isolated by multiple rounds of screens. The non-CTC BACs (RP11-340K9, CTD-2180L11, CTD-3155K1, and CTA-128F11) identified by sequence searches are denoted by a different font. STS markers from this physical map are available in GenBank (acc. nos. AQ939874–AQ939899). The *D5S2029* locus, a marker for the telomeric limit identified in primary leukemic blasts [4], is shown in bold. The centromeric limit *D5S1464* identified in ML3 cells is also in bold. The SP6 ends are denoted by arrowheads, and the T7 ends are blunt.

normal chromosome 5 and two marker chromosomes with chromosome 5 sequences, namely der(3) and der(5). While the der(3) contained 5q13.3–qter material juxtaposed to 3q in an inverted orientation, the der(5) is consistent with a 5q– chromosome. Yeast artificial chromosomes (YACs) spanning the inversion-coupled translocation interval in ML3 cells have been reported [2,4]. A doubly linked YAC tiling path that was contiguous across the *D5S1464–D5S620* interval delineated the chromosomal breakpoint at 5q13.3 [2,4]. Two nonoverlapping YACs (729F12 containing the marker *D5S1464*, and 965B11 containing the *D5S620* locus) hybridized to the two different derivative chromosomes, suggesting that the breakpoint resides between these YACs. However, two overlapping mega YACs, 940D1 and 934C2, also hybridized to the two different derivative chromosomes. We therefore hypothesized that the breakpoint resides within the small region of overlap between the YACs 940D1 and 934C2 and that the restricted sensitivity of FISH limited our ability to detect a split signal.

To refine further the physical map and delineate the breakpoint precisely, we constructed a BAC contig between the loci *D5S1464* and *D5S620*. Markers *D5S1464*, *D5S2029*, and *D5S620* were simultaneously screened against the Caltech CTC BAC library DNA pools. One clone, 448F14, was isolated for *D5S1464*, and the marker *D5S620* could be amplified from BACs 457H17 and 286G23. Clones 560O9, 421M10, and 236D20 were identified for the marker *D5S2029*. SP6 (arrow) and T7 (blunt) ends of these BACS were sequenced, and primers were designed and used again to screen the BAC DNA pools (Fig. 1). Of these, the SP6 end of 560O9 matched the marker *D5S2029* sequences. In addition, both ends of clone 457H17 and the T7 end of 560O9 could be localized within a fully sequenced CTA BAC clone 128F11 (GenBank acc. no. AC005406). Thus, we physically linked markers *D5S2029* and *D5S620*.

In the next step, partial genomic sequences of the BAC 560O9 that was available were also found to overlap with the end of the CTD BAC clone 2180L11 from the human genome survey (GenBank acc. no. AQ270315). Comparisons of partial genomic sequences from 2180L11 against the Genome Sequence Survey (GSS) database identified another CTD clone 3155K1 (GenBank acc. no. AQ786892). The end sequences for 2180L11 and 3155K1 were then used to design primers for another round of screening against the CTC BAC DNA pools to isolate BACs 360I18 and 484A7.

On the centromeric side, we had determined that the marker SHGC-34817 was telomeric of *D5S1464*, based on nonchimeric YACs. Therefore, this marker was used to isolate additional clones to fill the gap. This screen identified clones 448F14, 461G4, 518G22, 492D7, and 484A9. Of these, only 484A9 tested positive for the marker *D5S1740*. The T7 end of this clone was then used again to screen the BAC pool. Clones 209D19, 256E15, and 546C12 overlapped with the end sequence of 484A9, and all three clones also contained the marker *D5S1740*. To determine which clone extended most telomerically, the ends were sequenced and tested against each other by PCR. The T7 end of 546C12 did not amplify against any other clone and was then used in an additional round of screening to isolate BACs. The BAC 546I12 that was isolated by this screen contained the ends of both 3155K1 and 360I18.

The order of markers obtained by the BAC contig in Figure 1, which is triply linked for the most part, is consistent with the partial 1.47-Mb human genomic sequences (NT 007022) in the public domain. NT 007022 is a contig of assembled sequences from the Caltech C, D, and RP11 genomic libraries (Table 1). Four of the BAC clones isolated in our screen (CTC-564D18, CTC-448F14, CTC-484A9, and CTC-560O9) are also part of this assembly. Localization of the two markers *D5S1464* and *NIB1097* at nucleotides 126,661–126,926 and 971,391–971,606 of the NT 007022 sequences suggested the contig in Fig. 1 to be > 850 kb. At present, the NT 007022 sequences are annotated to encode 6 genes, with evidence for transcripts and 11 hypothetical genes predicted by genome-scanning methods (Table 1). It is interesting that genome scans did not detect any hypothetical transcripts in sequences from clones CTC-448F14, RP11-370B10, RP11-340K9, and CTC-484A9. The only gene localized between the nucleotides 53,299 and 399,399 is *SSBP2*, so designated because of its homology to a chicken gene, *CSDP*, that encodes a putative sequence-specific single-stranded DNA-binding activity, which turned out to be the target of disruption in ML3 cells.

Delineation of the 5q13.3 Breakpoint in ML3 Cells and Identification of a Candidate Gene

Selected individual BACs that localized to the region of overlap between the YACs 940D1 and 934C2 were used as FISH probes on ML3 cells. BACs 209D19 and 492D7, 448F14 and 564D18 were found to flank the breakpoint, whereas the BAC 484A9 gave a split signal, with a stronger signal on the der(3)

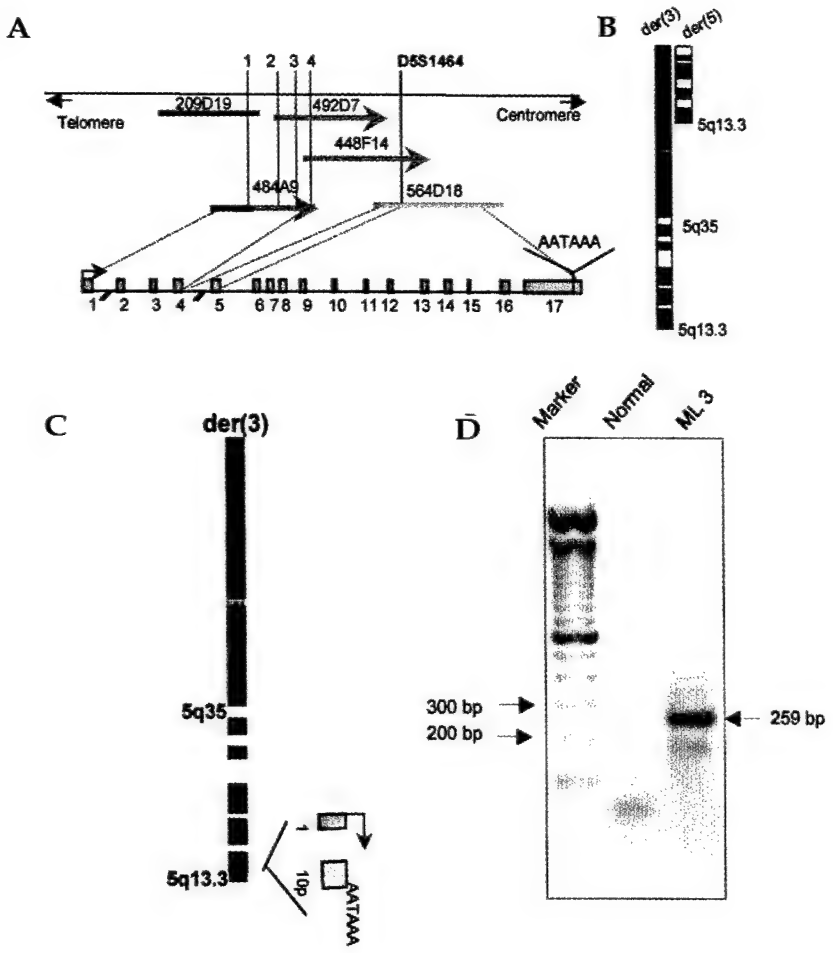

FIG. 2. The 5q13.3 chromosomal break in ML3 cells disrupts *SSBP2*. (A) FISH with BAC 209D19 sequences revealed hybridization to der(3), shown in red. Additionally, the BACs 546C12 and 360I18 (Fig. 1) also hybridized to der(3). BACs 492D7, 448F14, and 564D18 hybridized to der(5) chromosome shown in green. Draft sequences of BACs 484A9 and 564D18 allowed identification of 17 exons corresponding to the cDNA sequences. Splice donor and acceptor sequences flanked each exon, and a polyadenylation signal could be identified. GenBank NT_007022 sequences identified 16 of these exons between nucleotides 53,323 and 399,188. (B) Ideograms for the two derivative chromosomes from ML3 cells. Note that all the chromosome 5 sequences are retained, except for a disruption at 5q13.3. Localization of red and green hybridized BACs is indicated. (C) *SSBP2* is truncated within the first intron in ML3 cells. This illustration depicts the rearranged *SSBP2* allele on the der(3) chromosome in ML3 cells. The complex chromosomal changes at the telomere of the der(3) chromosome result in the transcription of the first exon of *SSBP2* that is spliced to a novel 3′-untranslated region (3′-UTR). The 3′-UTR with a polyadenylation signal matches genomic sequences from chromosome 10p15 with a flanking 3′ splice acceptor site. (D) *SSBP2* is truncated within the first intron in ML3 cells. Total RNA from ML3 cells was reverse-transcribed and amplified with an *SSBP2* exon 1–specific sense primer and an oligo-dT primer. The unique product was subcloned and sequenced, and primers were derived from the novel 3′-UTR sequences from chromosome 10p15 locus. The primer pair SSBP2 ATG.F, 5′-ATGTACGGCAAAGGCAAGAGT-3′, and 10p+195.R, 5′-GCACTTGTA GTCCCAACTACTC-3′, yielded the predicted 259-bp product from ML3 cells and not from normal mononuclear RNA. Sizes of the unique band and relevant molecular weight markers are denoted in base pairs.

tive 3′ exon of 277 bp with a polyadenylation signal (Fig. 2C). Therefore, the chromosomal break at 5q13.3 has occurred within the large first intron of *SSBP2*. The transcription potential of the novel 3′ exon was verified by the identification of flanking splice acceptor sites in the genomic sequences. The genomic sequences for the novel exon localized to chromosome 10p15.3. A review of the karyotype of ML3 cells revealed that this cell line is monosomic for chromosome 10. Thus, 10pter could have served as an ectopic telomere for the der(3) chromosome as a result of a cryptic unbalanced translocation with loss of chromosome 10 sequences. The authenticity of the RACE product was further confirmed by PCR with a combination of the *SSBP2*-specific forward primer and a reverse primer from the chromosome 10p exon. The anticipated product of 259 bp was detected in ML3 cells and not in normal mononuclear cells (Fig. 2D).

A Gene Family with High Evolutionary Conservation Revealed by the Deduced Open Reading Frame of *SSBP2*

Recursive searches of the databases yielded three identical full-length *SSBP2* cDNAs, one from normal $CD34^+$ hematopoietic stem cells (NM_012446), one from pituitary (AF077048), and one from fetal brain (AL080076). The predicted ORF of *SSBP2* shows similarity to the chicken *CSDP* [5]. In addition, the UniGene database annotates a ubiquitously expressed human *SSBP* gene (Hs. 266914), with high identity to *CSDP*. Hs 266914 sequences localize to chromosome 1p. We identified a third member of this family by recursive searches of dBEST. We obtained several IMAGE consortium cDNA clones and generated double-stranded sequences. The cDNA inserts were full length for *SSBP2*. However, for Hs. 266914, designated *SSBP3* (IMAGE: 1553027), and the third member, *SSBP4* (IMAGE: 2697135), the clones yielded partial sequences; exons 4–17 for *SSBP3* and exons 1–16 for *SSBP4*. We generated unique primers for the predicted exons 1–17 for both *SSBP3* and *SSBP4* and confirmed expression of both transcripts by RT-PCR (H.L. *et al.*, unpublished data). Alternative splicing as well as additional internal exons were identified for all three genes, suggesting the existence of multiple isoforms for each protein.

Thus, a total of three distinct but highly related ORFs were identified: 1) *SSBP3*, highest homology to the chicken

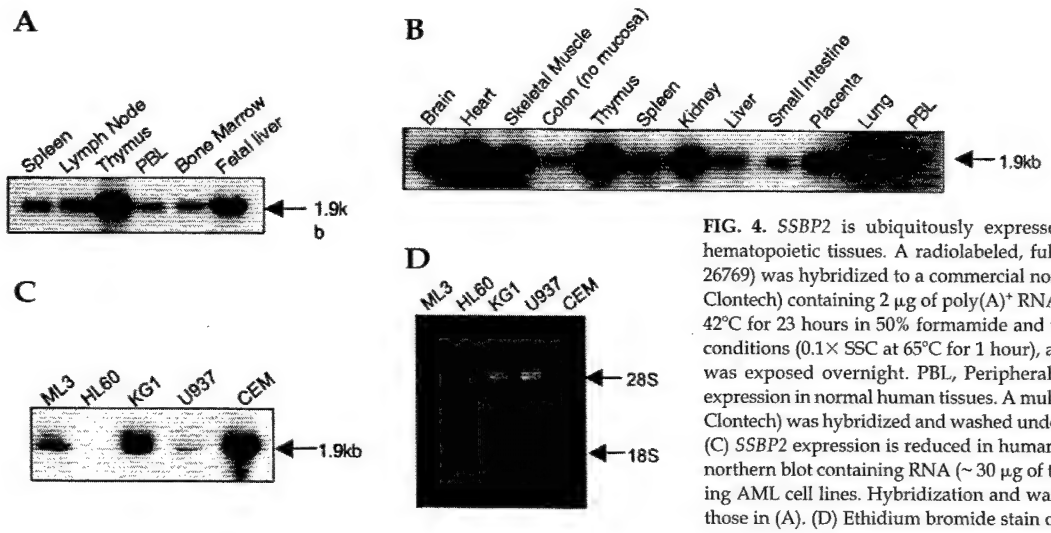

FIG. 4. *SSBP2* is ubiquitously expressed. (A) *SSBP2* expression in hematopoietic tissues. A radiolabeled, full-length cDNA clone (IMAGE 26769) was hybridized to a commercial northern blot (Immune system II; Clontech) containing 2 μg of poly(A)$^+$ RNA. Hybridizations were done at 42°C for 23 hours in 50% formamide and washed under high stringency conditions (0.1× SSC at 65°C for 1 hour), after which the autoradiograph was exposed overnight. PBL, Peripheral blood leukocytes. (B) *SSBP2* expression in normal human tissues. A multiple-tissue northern blot (MTS; Clontech) was hybridized and washed under the same conditions as in (A). (C) *SSBP2* expression is reduced in human AML cell lines as shown by a northern blot containing RNA (~ 30 μg of total) from exponentially growing AML cell lines. Hybridization and wash conditions were identical to those in (A). (D) Ethidium bromide stain of gel used in (C).

were high in heart, brain, kidney, and skeletal muscle (Fig. 4B).

We next examined the expression of *SSBP2* in leukemic cell lines (Fig. 4C). The childhood pre-T leukemia cell line CEM showed very high expression, analogous to the high expression seen in normal thymus. The AML cell line KG1 expressed *SSBP2* at significantly high levels. This is in sharp contrast to other AML cell lines (HL60, U937, and ML3), which showed reduced expression. The reduced transcript levels in ML3 are consistent with hemizygosity of *SSBP2*, because the truncated sequences (exons 2–17) from the der(5) chromosome do not appear to be transcribed in these cells.

Absence of Inactivating Mutations in *SSBP2*

To determine whether *SSBP2* is a classical tumor suppressor gene that functions by a recessive mechanism, we searched for intragenic mutations in the remaining allele in ML3 and HL60 cells, which are hemizygous for *SSBP2* (P.C. *et al.*, unpublished data), as well as in primary leukemic blasts from four patients (Table 2). The possibility of heterozygous mutations was also examined in KG1, TF1, and HEL cells. No gross inactivating mutations were detected in either the cell lines or leukemic blasts from the four patients, suggesting that *SSBP2* is not a common target of inactivating mutations. However, we noticed that the yields from the amplification reactions were low in all samples, suggesting *SSBP2* expression to be decreased in AML with the exception of KG1 cells.

Discussion

Loss of Gene Function by Chromosomal Translocations

Acquired complete and partial deletions of chromosomes 5 and 7 and trisomy 8 have long been recognized to be hallmarks of poor prognosis in myeloid malignancies [9,10]. Nonetheless, isolation of classical tumor suppressor genes that function by a recessive, two-hit mechanism from chromosomes 5 and 7 has remained elusive. The frequent co-segregation of two or more of these anomalies in the midst of other complexities has raised the possibility that the deletions in fact reflect an overall karyotypic instability. However, unbalanced translocation between chromosomes 5 and 17 causes loss of the wild-type allele of *TP53* despite the varied breakpoints on chromosome 17p [4,11,12]. Similarly, the *D5S672–D5S620* locus was hypothesized to harbor a critical gene that collaborates with loss of *TP53*, as this interval was invariantly deleted in patients or cell lines with loss of 17p.

Chromosome 5q13.3 is also a region of loss in hairy-cell leukemia (HCL), and three candidate genes have been identified from an interval characterized by constitutional inversion [13,14]. Although the interval delineated for HCL is close to *SSBP2*, it remains to be examined whether *SSBP2* is altered in HCL. We have hypothesized that the rare cases of AML with overlapping deletions at 5q13.3 due to unbalanced translocations pinpoint a critical suppressor element [4]. Notably, the ML3 cell line grossly retains all the 5q segments flanking 5q13.3, as revealed by heterozygosity for polymorphic loci and by FISH, thus excluding tumor suppressors from other chromosome 5 loci. The availability of this cell line provided a unique opportunity to isolate the *SSBP2* gene by a conventional positional cloning approach (Figs. 1 and 2).

TABLE 2: Absence of intragenic mutations in *SSBP2*[*]

Leukemia	Number of samples tested	Mutations
Primary AML/MDS	4	None
Cell lines	5 (ML3, KG1, HL60, TF1 and HEL)	None

[*]cDNA pools from primary leukemic blasts or cell lines were generated and screened for mutations by PCR with three pairs of primers. No mutations could be detected, although the PCR gave poor yields from the leukemic blasts and the ML3, HL60, TF1, and HEL cell lines.

Northern blot analysis. Northern blots containing poly(A)$^+$ RNA from hematopoietic tissues were purchased from Clontech Laboratories and probed with radiolabeled full-length *SSBP2* cDNA (IMAGE 26769). The blots were hybridized at 42°C for 23 hours in 50% formamide and washed under high-stringency conditions in 0.1× SSC at 65°C for 1 hour, after which the autoradiograph was exposed overnight.

Mutation screens. Total RNA (3 μg) from AML cell lines or leukemic blasts was reverse-transcribed with Moloney reverse transcriptase, and the cDNA pools were amplified by PCR with three pairs of nested, gene-specific primers. Two overlapping primer pairs (exon 1F+ exon7R and exon 4F+ stop29R) covered the entire coding region. The primer sequences are as follow: exon1F, 5′-GTTGACAGGTGCGTGACAGT-3′; exon7R, 5′-TCTGGGGAGTAATGGCTGAC-3′; exon4F, 5′-TGGAAGGCTTTTGCTTCACT-3′; stop29R, 5′-TGCAGTTCAGTTTAGGGCAAT-3′. Double-stranded sequences for both products, generated by the ABI automated sequencing system, were examined for intragenic mutations.

3′-RACE. 3′-RACE was conducted on cDNA pools from RNA reverse-transcribed with oligo-dT, essentially as described [33].

Acknowledgments

We thank Xiuying Lina Wu and Lisa Chu for assistance with FISH analysis; Rashmi Pershad for the automated DNA sequence analysis; Walter Hittelman for comments on the manuscript; Elva Lopez for assistance with manuscript preparation; and Jan-Fang Cheng (Joint Genome Institute, Walnut Creek, CA) for help with the BAC contig. This work was supported by the Department of Defense (DAMD 17-99-1-9267), National Institutes of Health, and funds from Ladies Leukemia League (Metairie, LA) and Abraham and Phyllis J. and Phyllis Katz Foundation to L.N. The automated DNA sequencing facility is supported by core grant CA16672 to the University of Texas M. D. Anderson Cancer Center.

RECEIVED FOR PUBLICATION JANUARY 29; ACCEPTED MAY 9, 2002.

References

1. Pedersen, B. (1996). Anatomy of the 5q- deletion: different sex ratios and deleted 5q bands in MDS and AML. *Leukemia* **10:**1883-1890.
2. Fairman, J., *et al.* (1996). Translocations and deletions of 5q13.1 in myelodysplasia and acute myelogenous leukemia: evidence for a novel critical locus. *Blood* **88:** 2259–2266.
3. Castro, P. D., Fairman, J., and Nagarajan, L. (1998). The unexplored 5q13 locus: a role in hematopoietic malignancies. *Leuk. Lymphoma* **30:** 443–448.
4. Castro, P., Liang, J. C., and Nagarajan, L. (2000). Deletions of chromosome 5q13.3 and 17p loci cooperate in myeloid neoplasms. *Blood* **95:** 2138–2143.
5. Bayarsaihan, D., Soto, R. J., and Lukens, L. N. (1998). Cloning and characterization of a novel sequence-specific single-stranded DNA-binding protein. *Biochem. J.* **331:** 447–452.
6. Conner, J., and Liu, Z. (2000). LEUNIG, a putative transcriptional corepressor that regulates AGAMOUS expression during flower development. *Proc. Natl. Acad. Sci. USA* **97:** 12902–12907.
7. Kobayashi, O., Yoshimoto, H., and Sone, H. (1999). Analysis of the genes activated by the *FLO8* gene in *Saccharomyces cerevisiae*. *Curr. Genet.* **36:** 256–261.
8. Gomes, X. V., and Wold, M. S. (1996). Functional domains of the 70-kilodalton subunit of human replication protein A. *Biochemistry* **35:** 10558–10568.
9. Jacobs, R. H., *et al.* (1986). Prognostic implications of morphology and karyotype in primary myelodysplastic syndromes. *Blood* **67:** 1765–1772.
10. Estey, E. H., *et al.* (1987). Karyotype is prognostically more important than the FAB system's distinction between myelodysplastic syndrome and acute myelogenous leukemia. *Hematol. Pathol.* **1:** 203–208.
11. Lai, J. L., *et al.* (1995). Myelodysplastic syndromes and acute myeloid leukemia with 17p deletion. An entity characterized by specific dysgranulopoiesis and a high incidence of *p53* mutations. *Leukemia* **9:** 370–381.
12. Soenen, V., *et al.* (1998). 17p Deletion in acute myeloid leukemia and myelodysplastic syndrome. Analysis of breakpoints and deleted segments by fluorescence *in situ*. *Blood* **91:** 1008–1015.
13. Wu, X., *et al.* (1997) Characterization of a hairy cell leukemia-associated 5q13.3 inversion breakpoint. *Genes Chromosomes Cancer* **20:** 337–346.
14. Wu, X., *et al.* (1999). Molecular analysis of the human chromosome 5q13.3 region in patients with hairy cell leukemia and identification of tumor suppressor gene candidates. *Genomics* **60:** 161–171.
15. Mu, Z. M., Chin, K. V., Liu, J. H., Lozano, G., and Chang, K. S. (1994). PML, a growth suppressor disrupted in acute promyelocytic leukemia. *Mol. Cell. Biol.* **10:** 6858–6867.
16. Guo, A., *et al.* (2000). The function of PML in *p53*-dependent apoptosis. *Nat. Cell Biol.* **10:** 730–736.
17. Dobson, C. L., Warren, A. J., Pannell, R., Forster, A., and Rabbitts, T. H. (2000). Tumorigenesis in mice with a fusion of the leukaemia oncogene Mll and the bacterial *lacZ* gene. *EMBO J.* **19:** 843–851.
18. Song, W. J., *et al.* (1999). Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat. Genet.* **23:** 166–175.
19. Imai, Y., *et al.* (2000). Mutations of the *AML1* gene in myelodysplastic syndrome and their functional implications in leukemogenesis. *Blood* **96:** 3154–3160.
20. Osato, M., *et al.* (1999). Ito Biallelic and heterozygous point mutations in the runt domain of the *AML1/PEBP2α* gene associated with myeloblastic leukemias. *Blood* **93:** 1817–1824.
21. Veldman, T., Vignon, C., Schrock, E., Rowley, J. D., and Ried, T. (1997). Hidden chromosome abnormalities in haematological malignancies detected by multicolour spectral karyotyping. *Nat. Genet.* **4:** 406–410.
22. Ning, Y., Liang, J. C., Nagarajan, L., Schrock, E., and Ried, T. (1998). Characterization of 5q deletions by subtelomeric probes and spectral karyotyping. *Cancer Genet. Cytogenet.* **103:** 170–172.
23. Mirkin, S. M., and Frank-Kamenetskii, M. D. (1994). H-DNA and related structures. *Annu. Rev. Biophys. Biomol. Struct.* **23:** 541–576.
24. Darbinian, N., *et al.* (2001). Growth inhibition of glioblastoma cells by human Pur (α). *J. Cell. Physiol.* **189:** 334–340
25. Gallia, G. L., Johnson, E. M., and Khalili, K. (2000). Purα: a multifunctional single-stranded DNA- and RNA-binding protein. *Nucleic Acids Res.* **28:** 3197–3205
26. Arlt, M. F., Li, M., Herzog, T. J., and Goodfellow, P. J. (1999). A 1-Mb bacterial clone contig spanning the endometrial cancer deletion region at 1p32–p33. *Genomics* **57:** 62–69.
27. Peng, H., *et al.* (2000). ARHI is the center of allelic deletion on chromosome 1p31 in ovarian and breast cancers. *Int. J. Cancer* **86:** 690–694.
28. Avigad, S., *et al.* (1997). Prognostic relevance of genetic alterations in the p32 region of chromosome1 in neuroblastoma. *Eur. J. Cancer* **33:** 1983–1985.
29. Mathew, S., Murty, V. V., Bosl, G. J., and Chaganti, R. S. (1994). Loss of heterozygosity identifies multiple sites of allelic deletions on chromosome 1 in human male germ cell tumors. *Cancer Res.* **54:** 6265–6269
30. Gasparian, A. V., *et al.* (1998). Allelic imbalance and instability of microsatellite loci on chromosome 1p in human non-small-cell lung cancer. *Br. J. Cancer* **77:** 1604–1611.
31. Sanchez-Cespedes, M., *et al.* (2001). Chromosomal alterations in lung adenocarcinoma from smokers and nonsmokers. *Cancer Res.* **61:** 1309–1313.
32. Zhao, L., Van Oort, J., Cork, A., and Liang, J. C. (1993). Comparison between interphase and metaphase cytogenetics in detecting chromosome 7 defects in hematological neoplasias. *Am. J. Hematol.* **43:** 205–211.
33. Frohman, M. A. (1990). RACE: rapid amplification of cDNA ends. In *PCR Protocols: A Guide to Methods and Applications* (M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White, Eds.), pp. 28–38. Academic Press, San Diego, CA.

A human Mix-like homeobox gene *MIXL* shows functional similarity to *Xenopus* Mix.1

Wei Guo, Agnes Pui-yee Chan, Hong Liang, Eric D. Wieder, Jeffrey J. Molldrem, Laurence D. Etkin, and Lalitha Nagarajan

Molecular events involved in specification of early hematopoietic system are not well known. In *Xenopus*, a paired-box homeodomain family (Mix.1-4) has been implicated in this process. Although Mix-like homeobox genes have been isolated from chicken (*CMIX*) and mice (*Mml/MIXL1*), isolation of a human Mix-like gene has remained elusive. We have recently isolated and characterized a novel human Mix-like homeobox gene with a predicted open reading frame of 232 amino acids designated the Mix.1 homeobox (*Xenopus laevis*)–like gene (*MIXL*). The overall identity of this novel protein to CMIX and Mml/MIXL1 is 41% and 69%, respectively. However, the identity in the homeodomain is 66% to that of *Xenopus* Mix.1, 79% to that of CMIX, and 94% to that of Mml/MIXL1. In normal hematopoiesis, *MIXL* expression appears to be restricted to immature B- and T-lymphoid cells. Several acute leukemic cell lines of B, T, and myeloid lineage express MIXL suggesting a survival/block in differentiation advantage. Furthermore, *Xenopus* animal cap assay revealed that MIXL could induce expression of the α-globin gene, suggesting a functional conservation of the homeodomain. Isolation of the *MIXL* gene is the first step toward understanding novel regulatory circuits in early hematopoietic differentiation and malignant transformation. (Blood. 2002;100:89-95)

Introduction

Mechanisms governing specification, proliferation, and differentiation of hematopoietic progenitors into terminally differentiated circulating elements have begun to unfold in the past decade. Several transcription factors, uncovered by specific chromosomal translocations, are also regulators of normal hematopoietic differentiation as demonstrated by loss-of-function studies in murine models. More importantly, the striking evolutionary conservation of lineage-commitment mechanisms between fruit flies and mammals[1] suggests that studies on mammalian orthologs of genes that regulate early hematopoiesis in invertebrates or amphibians may shed light on novel regulatory circuits.

Mix.1, a pairedlike homeobox gene in *Xenopus*, was initially identified as an inducer of ventral mesoderm and/or endoderm.[2,3] Subsequently, several Mix.1-like genes were isolated and shown to be involved in the regulation of mesoderm and/or endoderm formation.[4-7] At present, the Mix.1-like gene family includes the Mix family (Mix.1-4), Bix family (Bix1-4), and Mixer in *Xenopus*.[3-7] However, in chicken (*CMIX*), mice (*Mml*) and humans, the Mix-like homeobox genes appear to be single copies.[8-11] Although the proteins encoded by Mix-like genes vary in size, they are all modular with a highly conserved paired-type homeodomain and a conserved carboxy-terminal acidic domain.

The *Xenopus* Mix.1 gene is implicated in the process of patterning ventral mesoderm to hematopoietic fate induced by bone morphogenetic protein 4 (BMP-4), a member of the transforming growth factor β (TGF-β) family.[12] In combination with mesoderm inducers fibroblast growth factor (FGF) or activin, BMP-4 induces abundant hematopoietic mesoderm in *Xenopus* animal caps.[13,14] A dominant-negative mutant of Mix.1 blocks the BMP-4 response in this system.[12] Consistent with the *Xenopus* studies, BMP-4 expression in human embryos is polarized to the ventral wall of the dorsal aorta, which is associated with clusters of hematopoietic cells in the aorta-gonad-mesonephros (AGM) region.[15] Strong BMP-4 expression is also detected in cells specifically associated with blood islands in human embryonic yolk sac.[15] These observations in human embryos suggest a role for BMP-4 in the initiation of both yolk sac and AGM hematopoiesis.

The *SMAD5* gene, a signal transducer of BMP-4, localizes to human chromosome 5q31.1, a region of loss in human acute myelogenous leukemia (AML).[16] Our previous studies characterized a carboxy-terminally truncated alternate splice form *SMAD5*β that is preferentially expressed in hematopoietic stem cells and leukemia.[17] Because of our interests in the functional consequences of persistent *SMAD5*β expression and possible haploinsufficiency for full-length *SMAD5*, we searched for a putative downstream target, the human Mix-like homeobox gene. Here, we report the isolation and characterization of the human *MIXL* (Mix.1 homeobox [*Xenopus laevis*]-like) gene, which mimics *Xenopus* Mix.1 in functional assays. Our findings raise the possibility of an essential role for *MIXL* in embryonic and adult hematopoiesis.

Materials and methods

Databases and bioinformatics

The databases Nucleotide (NR), High Throughput Genomic Sequences (HTGS), and Expressed Sequence Tags (dbEST) at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov), as well as the

From the Department of Molecular Genetics and the Department of Blood and Marrow Transplantation, University of Texas M. D. Anderson Cancer Center, Houston.

Submitted September 10, 2001; accepted February 14, 2002.

Supported by the Department of Defense (DAMD 17-99-1-9267), National Institutes of Health (CA66982 and CA55164), funds from Katz Foundation (to L.N.), and core grant CA16672 to the University of Texas M. D. Anderson Cancer Center.

Reprints: Lalitha Nagarajan, Department of Molecular Genetics, University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd, Box 45, Houston, TX 77030; e-mail: lalitha@odin.mdacc.tmc.edu.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 U.S.C. section 1734.

chromosome 1 sequences at the Sanger Center (http://www.sanger.ac.uk/HGP/Chr1/), were used to search for complementary DNAs (cDNAs), sequence tagged sites (STSs), and genomic sequences. High-resolution fluorescence in situ hybridization (FISH) of chromosome 1 markers was accessed at the website http://cgap.nci.nih.gov/Chromosomes/BAC_Clone_Map?CHR = 1. The Mitelman Database of Chromosome Aberrations in Cancer culled from the literature by F. Mitelman, B. Johansson, and F. Mertens (editors) was accessed at the Cancer Genome Anatomy Project (CGAP) web site http://cgap.nci.nih.gov/chromosomes/Mitelman.

Expressed sequence tags and bacterial artificial chromosome

The EST clones AA847809, AA911377, and AI654861 and the bacterial artificial chromosome (BAC) clone CITB-HSP-C 216I11 were purchased from Research Genetics (Huntsville, AL).

Polymerase chain reaction

All the polymerase chain reactions (PCRs) were conducted under the following conditions: an initial denaturation for 4 minutes at 95°C, followed by 35 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C to 60°C for 30 seconds, and extension at 72°C for 30 seconds, plus a final extension for 10 minutes at 72°C. The primers were used at the concentrations of 10 pmol in a total volume of 20 μL. The *Taq* polymerase was obtained from Gibco BRL (Rockville, MD).

The CITB Human BAC DNA Pool (CITB-HSP-C Library, Research Genetics) was screened with a pair of *MIXL*-specific primers: *MIXL*-EST1 forward primer 5′-CAA AGC TGG ACT CAT GGG AG-3′ and reverse primer 5′-AAC GAA TGC GGG AAC TCT GG-3′. The *MIXL* intron was amplified from BAC 216I11 with the Expand High Fidelity PCR System (Roche, Indianapolis, IN). The primer pair used to amplify the putative intronic sequences was *MIXL*-HD forward primer 5′-TTT CAG CGC CGA ACA G-3′ and reverse primer 5′-ATC TCC GGC CTA GCC AAA GG-3′. The 3′ rapid amplification of cDNA ends (RACE) was performed with the primer MIX-3RACE 5′-GCG TGC CAA GTC TCG GCG TCA G-3′.

Radiation hybrid mapping

The G3 radiation hybrid (RH) panel (a panel of 83 whole genome human-hamster radiation hybrids) was used to generate the RH linkage data.[19] Eighty-three PCR reactions were performed with a pair of *MIXL* primers: *MIXL*-EST2 forward primer 5′-TCT GGG AGA AAT CCG GAT AAG C-3′ and reverse primer 5′-TGT GAG AGG TGC TGT CAA AAC C-3′. The data were confirmed by PCR with a second pair of primers (*MIXL*-EST1). The combined RH linkage data set where 0 represents negative signals and 1 represents positive signals, was submitted to the Stanford Radiation Hybrid Mapping server (http://shgc.stanford.edu/RH/rhserverformnew.html) for statistical linkage analysis.

Northern blotting

The Human Immune System Multi-tissue Northern Blot II was purchased from Clontech (Palo Alto, CA). The blot was prehybridized for 5 hours with the prehybridization solution (5 times SSPE (150 mM sodium chloride, 10 mM sodium phosphate, 18 mM ethylenediaminetetraacetic acid [EDTA]), 10 times Denhardt solution, 2.0% sodium dodecyl sulfate (SDS), 50% deionized formamide, 100 μg/mL sheared salmon sperm DNA). The plasmid pCR2.1-TOPO-MIXL containing a 1 kilobase-pair (Kbp) *MIXL* insert (424 bp coding sequences including the Mix-like homeobox and 543 bp 3′ untranslated region [UTR] sequences) in the vector pCR2.1-TOPO (Invitrogen, Carlsbad, CA) was randomly primed with [α-^{32}P]dCTP (Random Primed DNA Labeling Kit from Roche) and added to the hybridization solution at a final concentration of 1.3×10^7 cpm/mL. The blot was rinsed at room temperature 3 times with wash solution I (2 times SSC, 0.1% SDS) after 28 hours of hybridization. Final washes were done at 55°C with wash solution II (0.5 times SSC, 0.1% SDS) for 30 minutes before exposing the blot to x-ray films (X-OMT; Kodak, New Haven, CT) for 5 days. The blot was rehybridized with human β-actin cDNA (Clontech) to evaluate messenger RNA (mRNA) loading.

Cell sorting

Previously cryopreserved normal donor bone marrow (6×10^7 cells) or peripheral blood mononuclear cells (1×10^7 cells) were thawed, resuspended at 2×10^7 cells/mL in phosphate-buffered saline (PBS) and stained with appropriate combinations of monoclonal antibodies for 30 minutes at 4°C. The following monoclonal antibodies were used: CD5–fluorescein isothiocyanate (FITC), CD13-phycoerythrin (PE; Becton Dickinson, San Jose, CA); CD34-ECD (phycoerythrin-Texas Red-x; Beckman/Coulter, Hileah, FL); glycophorin A FITC, CD3-allophycocyanin (APC), CD45-APC-cyanine7 (Cy7), CD19-tri-color (R-PE-Cy5 tandem) (TC) (Caltag Labs, Burlingame, CA). After staining, cells were washed and resuspended for subsequent analysis and cell sorting on a MoFlo cell sorter (Cytomation, Fort Collins, CO). For bone marrow, live cells ($CD45^+$) from bone marrow were gated for CD34 expression. Cells expressing CD34 (5.77% of $CD45^+$ cells) were subgated for myeloid, T, and B lymphoid progenitors. Myeloid ($CD34^+CD45^+CD13^+$), T ($CD34^+CD45^+CD5^+$), or B ($CD34^+CD45^+CD19^+$) progenitors were sorted and deposited directly into tubes containing TriZol (Gibco). For peripheral blood, subgroups of cells were identified based on scatter characteristics and antibody staining: $CD13^+$ monocytes, glycophorin A^+ erythrocytes, $CD19^+$ B lymphocytes, or $CD3^+$ T lymphocytes were sorted for total RNA extraction.

Reverse transcriptase-coupled PCR

Total RNAs from human cells, *Xenopus* embryos, or animal caps were extracted with TRIzol (Gibco) according to the manufacturer's instructions. The reverse transcription was performed with the oligo dT (Gibco) or the primer Qt.[18] The SuperScript II RNase H^- M-MLV reverse transcriptase was purchased from Gibco. After reverse transcription (RT), the reactions were diluted 1:40, and 2 μL aliquots were used as templates for PCR reactions. *MIXL* expression in sorted cells was determined by RT-PCR with the *MIXL-EXP* primers 5′-GGTACCCCGACATCCACTTG-3′ and 5′-CTCCCATGAGTCCAGCTTTG-3′ and the β-actin primers 5′-AGAGCAAGAGAGGCATCCTC-3′ and 5′-ATAGCACAGCCTGGATAGCA-3′, followed by Southern blotting with a *MIXL* cDNA fragment and a β-actin cDNA fragment as described below. For each sample, 5 parallel reactions of RT-PCR were performed with different PCR cycles (18, 21, 24, 27, or 30 cycles).

The RT-PCR products were resolved on 1.5% agarose gels and then transferred to the Hybond N+ membranes (Amersham, Piscataway, NJ) with 0.4 M sodium hydroxide. The DNA on the blots was cross-linked by UV (Hoefer UV cross-linker, Amersham). The blots were prehybridized for 1 hour with the prehybridization solution (5 times SSPE, 5 times Denhardt solution, 0.5% SDS, 100 μg/mL sheared salmon sperm DNA). The 1 kbp *MIXL* and 286 bp β-actin cDNA fragments were randomly primed with [α-^{32}P]dCTP (Random Primed DNA Labeling Kit from Roche) and added to the hybridization solution at a final concentration of 2.6×10^6 cpm/mL. After 14 hours of hybridization, the blot was rinsed at room temperature 3 times with wash solution I (2 times SSC, 0.1% SDS). Final washes were done at 65°C with wash solution II (0.05-0.1 times SSC, 0.1% SDS) for 30 minutes before exposing the blot to the Phosphor Imager screen (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) for 1 hour.

Generation of antisera and immunoblotting

The amino-terminal antigen (TAESRALQFAEGAAF; single-letter amino acid codes) and carboxy-terminal antigen (GSKLDSWEEHIFSAF) from the predicted *MIXL* open reading frame (ORF) were synthesized in the synthetic antigen core facility at our institution. Rabbit polyclonal antibodies/antisera (anti-MIXL-N and anti-MIXL-C) were raised against both the peptides (Bethyl Lab, Montgomery, TX). The anti-*MIXL*-N antiserum was further affinity-purified (Bethyl Lab). Both antibodies were evaluated for specificity in nuclear extracts of transiently transfected COS-1 or 293T cells. Although both antibodies detected the same specific band, the affinity-purified amino-terminal antibody anti-*MIXL*-N with a higher titer was used in further studies.

Nuclear extracts were prepared by lysing 4×10^6 cells (rinsed twice with 1 times PBS) in 400 μL lysis buffer (10 mM HEPES [4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid], pH 7.9, 10 mM KCl, 0.1 mM EDTA,

0.1 mM ethyleneglycotetraacetic acid (EGTA), 1 mM dithiothreitol [DTT], 1 mM phenylmethylsylfonyl fluoride [PMSF], 2 μg/mL leupeptin, 2 μg/mL pepstatin A, 2 μg/mL aprotinin, 500 μg/mL benzamidine, 1 mM Na_3VO_4). After 15 minutes of lysis on ice, 12.5 μL of 10% Nonidet P-40 (NP-40) was added to the lysates and vortexed for 15 seconds. The pellets obtained after a quick spin at 14 000 rpm were resuspended in extraction buffer (20 mM HEPES, pH 7.9, 0.4 M NaCl, 1 mM EDTA, 1 mM EGTA, 1 mM DTT, 1 mM PMSF, 2 μg/mL leupeptin, 2 μg/mL pepstatin A, 2 μg/mL aprotinin, 500 μg/mL benzamidine, 1 mM Na_3VO_4). The extraction was continued on ice for 30 minutes with intermittent vortexing. The supernatant nuclear proteins obtained after centrifugation at 14 000 rpm for 5 minutes were resolved (50 μg proteins/lane) on precast 10% SDS-polyacrylamide gels (Invitrogen). After electrophoresis, the proteins were transferred to Hybond P nylon membrane (Amersham) at 30 V overnight. The protocol for immunoblotting was essentially as detailed elsewhere.[20] The mouse monoclonal antibody against human histone H1 (Santa Cruz Biotechnology, Santa Cruz, CA) was used as loading control.

In vitro transcription, embryo injection, and animal cap assay

The predicted ORF of *MIXL* was cloned into the vector pBluescript RN3 (a kind gift from Nigel Garrett and Patrick Lemaire) and designated pRN3-MIXL. The mutated *MIXL* with an in-frame deletion of the entire homeodomain was also cloned into the same vector and designated pRN3-ΔMIXL. Capped mRNAs of *MIXL* and truncated *MIXL* were synthesized from the *Sfi*I-linerized constructs pRN3-MIXL and pRN3-ΔMIXL by using T3 RNA polymerase (mMESSAGE mMACHINE; Ambion, Austin, TX), and purified through G50 Sephadex Column (Roche).

For animal cap assay, *Xenopus* embryos were in vitro fertilized, dejellied, and cultivated, and injected with synthetic mRNA at the one-cell stage in the animal pole. Then, animal caps were explanted at stage 8 and cultured to the sibling stage 36 in the presence of 200 ng/mL recombinant human basic FGF (bFGF). The expression of *Xenopus* αT4-globin gene was examined by RT-PCR with the primer pair 5′-TTG CTG TCT CAC ACC ATC CAG G-3′ and 5′-TCT GTA CTT GGA GGT GAG GAC G-3′. EF1α, an endoderm marker 5′-GGA AAG TCC ACA ACA ACT GG-3′ and 5′-GGA GCA TCA ATG ATA GTG AC-3′, was used as control. The multiplex PCR of 27 cycles was performed under conditions described above.

Results

Isolation of human *MIXL*

To molecularly clone the human Mix.1-like gene(s), we searched the human EST database against the homeodomain sequences of *Xenopus* Mix.1 with the tBLASTN software program (http://www.ncbi.nlm.nih.gov/BLAST) in the public domain. A single EST clone AA847809, containing a partial homeodomain with 66% identity and 72% similarity in the predicted ORF to that of *Xenopus* Mix.1, was identified from a $CD20^+$ germinal center pre-B cell library. Double-stranded sequencing of the 1.8 kbp insert of AA847809 allowed us to deduce a partial ORF of 155 amino acids followed by 3′ UTR of 1307 nucleotides, including 214 bp of an alu repeat and no polydenylation signal. Surprisingly, recursive searches with the AA847809 sequences identified only 3 other clones: AA911377, from a mammary carcinoma cDNA library, and AI654861 and BE552088, from germ cell tumors, in the entire human EST database of about 3.7 million entries from more than 30 different tissues. Thus the expression of the Mix.1-like homeobox gene may be tightly regulated.

Because all the 4 clones contain partial 3′ coding sequences of the presumptive gene, we attempted to isolate the 5′ and 3′ sequences by RACE of cDNA pools from fetal brain. Unfortunately, this approach did not yield readily discernible 5′ products due to the high GC content (> 75%) of the 5′ end, whereas the 3′-RACE yielded a 1.3-kbp product (data not shown).

An alternate approach was to isolate BAC clones, because these could be sequenced to deduce additional 5′ coding sequences. A pair of primers (MIXL-EST1) covering the unique 3′ region from the carboxy-terminal domain to the 3′ UTR yielded the predicted 271-bp fragment from normal human genomic DNA. This primer pair was used to screen the CITB-HSP-C human BAC library, and BAC clone 216I11 was isolated. By assembling sequences generated from the BAC clone with *MIXL*-specific primer with the sequences from EST clone AA847809, we predicted an ORF of 232 amino acids, similar to the previously reported murine Mix-like (*Mml*/*MIXL1*) gene. We designated the gene *MIXL* (Mix.1 homeobox [*Xenopus laevis*]-like gene, GenBank accession no. AF211891) in consultation with the Human Genome Nomenclature Committee.

The deduced ORF has a putative ATG start codon with a good Kozak motif (G_{-3} and G_{+4} [GGAGCGatgG])[21] and contains a proline-rich amino-terminal region, a pairedlike homeodomain, and a carboxy-terminal acidic domain (Figure 1A).

To characterize the genomic organization of *MIXL*, we designed several pairs of primers to probe the exon-intron junctions in BAC 216I11 by PCR. By using long-range PCR with the primer pair MIXL-HD that spans the codons 92-157, we amplified a 1.6-kbp genomic fragment. As a result, a 1.4-kbp intron was identified in the middle of the homeobox (Figure 1B). The possibility of one or more untranslated 5′ exons cannot be excluded at the present time.

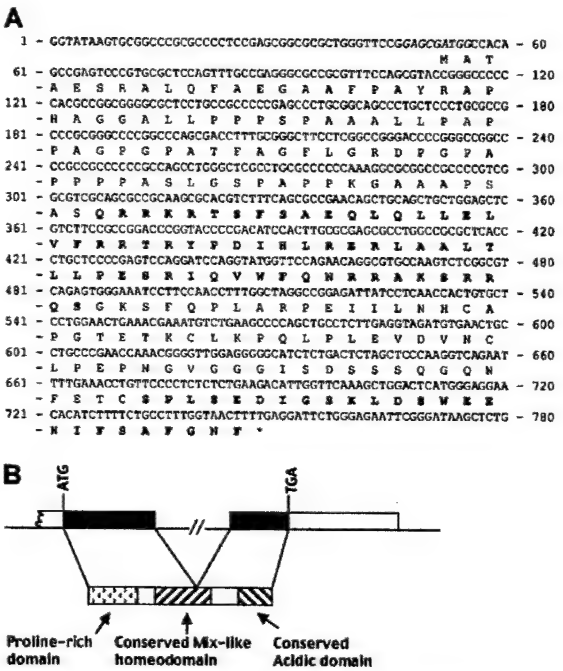

Figure 1. Human Mix.1-like homeobox gene *MIXL*. (A) The predicted ORF of *MIXL*. Nucleotides 1-285 were derived from sequencing BAC CITB-HSP-C216I11. The rest was from cDNA clones AA847809, AA911377, and AI654861. The Kozak consensus around the initiation Met are in italics. The conserved Mix-like homeodomain and carboxy-terminal domain are in bold. (B) Genomic organization of *MIXL*. The 1.6-kbp genomic fragment containing the intron was amplified from BAC 216I11 with the primer pair *MIXL*-HD, subcloned, and sequenced. The exon-intron junctions with perfect splice donor and acceptor consensus thus identified between nucleotides 448 and 449 (panel A) were confirmed by direct sequencing of BAC 216I11. Exon 1 is predicted to be more than 393 bp and exon 2 to be about 1.3 kbp. The protein domains of the predicted *MIXL* ORF are indicated. The amino-terminal proline-rich domain, not found in *Xenopus* Mix.1, is unique to *MIXL* and *Mml*.

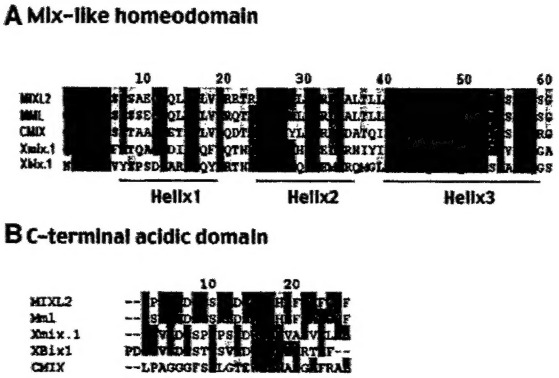

Figure 2. High degree of conservation in the homeo- and acidic carboxy-terminal domains in evolution. (A) ClustalW alignment of the 60 amino acid homeodomain sequences. The MIXL homeodomain shows 66% identity to that of *Xenopus* Mix.1. Residues constituting the 3 helical motifs are denoted. Note the high identity in the amino-terminal arm and helix 3 responsible for DNA binding. (B) Homology in the carboxy-terminal acidic domain. The highly conserved SxxSD/E and WEE motifs are denoted.

Both Mix-like homeodomain and carboxy-terminal domain are evolutionarily conserved

To determine the homology and phylogenetic distances between the human *MIXL*, other Mix-like genes (*Mml* and *CMIX*), and *Xenopus* Mix family members, the homeodomains of these genes were compared by the ClustalW program.[22] A neighbor joining phylogenetic tree revealed the human *MIXL* gene to be more closely related to the mouse *Mml*/*MIXL*1 gene than other genes (data not shown). The 2 proteins showed an overall identity of 69% with a 94% identity in the homeodomain (Figure 2A). Furthermore, the *MIXL* homeodomain is closer to that of *Xenopus* Mix.1 with 66% identity and 72% similarity than to that of *Xenopus* Bix.1 at 61% identity and 69% similarity (Figure 2A).

In the carboxy-terminal domain, MIXL also shows homology to Xenopus Mix.1 and other Mix-like proteins. As shown in Figure 2B, the 2 motifs SxxSD/E and WEE in the domain are conserved among most Mix-like proteins, suggesting that the carboxy-terminal domain is essential for *MIXL* function.

Localization of *MIXL* to chromosome 1q42.1

We localized *MIXL* by RH mapping to examine whether the gene maps to a chromosomal region implicated in malignancies. An RH data set generated by PCR with *MIXL* primers was submitted to the Stanford RH server for linkage mapping. The statistical analysis then demonstrated that the *MIXL* gene is linked to several markers (LOD > 6) located on chromosome 1q42, including SHGC-355541, SHGC-30345, SHGC-30224, and SHGC-37008 (Figure 3). FISH studies with BAC 216I11 further confirmed the mapping to chromosome 1q42 (J. Ma, W.G., and L.N., unpublished results, January, 2000).

To further refine the mapping, a small BAC contig was built *in silico* by a search of the human genome database (Figure 3). *MIXL* sequences were detected in BACs RP11-257I14 (accession no. AL592045) and RP11-588H15 (accession no. AC021883). The RP11-588H15 sequences overlap with the RP11-15H13 (accession no. AC011651), which contains SHGC-30345, a marker identified by the RH mapping. Furthermore, BAC RP11-257I14 overlaps with BAC RP11-396C23 (accession no. AL512343) containing the polymorphic marker *D1S479*, which is physically localized to the human chromosome 1q42.12 by high-resolution FISH mapping in the public domain. Therefore, the *MIXL* gene is physically localized to human chromosome 1q42.1, a region with increased copy number (gain) as detected by comparative genomic hybridization in mammary cancer and lymphoid malignancies.[23-25]

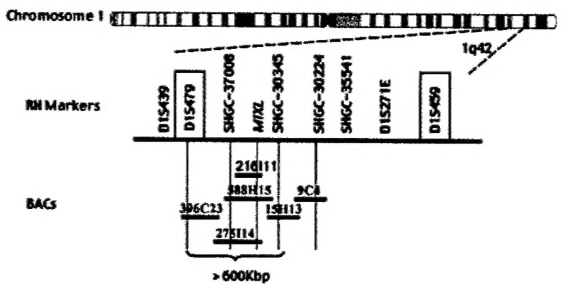

Figure 3. *MIXL* localizes to human chromosome 1q42. *MIXL* was localized to human chromosome 1q42 and linked to the markers in bold by radiation hybrid mapping. Tentative genomic sequences from BACs 216I11, RP11-588H15, and RP11-275I14 contain the *MIXL* gene. BACs 588H15 and 275I14 also span the marker SHGC-37008. BAC RP11-15H13, which overlaps with BAC 588H15, contains the marker SHGC-30345. BAC RP11-396C23 overlapping with BAC 275I14 contains the microsatellite marker *D1S479*. The microsatellite markers *D1S479* and *D1S459* (in text box) localize to band 1q42.12 and 1q42.2b by high-resolution FISH.

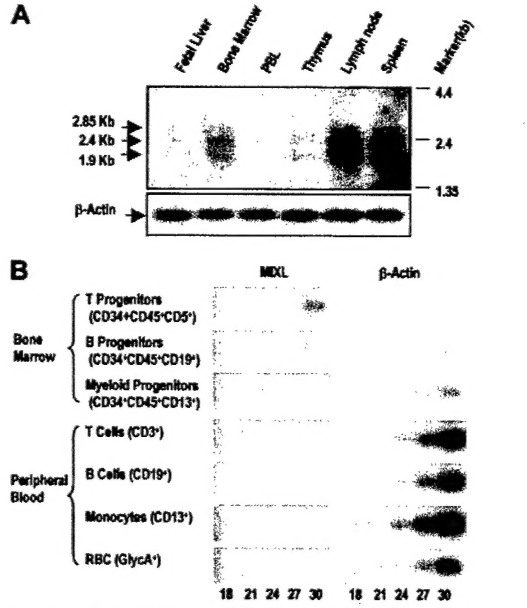

Figure 4. *MIXL* expression is restricted in normal hematopoietic tissues. (A) Human immune system multitissue Northern blot (Clontech Laboratories) was hybridized overnight with radiolabeled *MIXL* cDNA probe as described in "Materials and methods," and the blot was exposed to x-ray film for 5 days. Arrows denote the 3 bands detected. The autoradiograph of the same blot rehybridized with β-actin is shown in the lower panel. PBL indicates peripheral blood leukocytes. (B) The T, B, or myeloid progenitors (1×10^3, 2.5×10^3, and 9×10^3 cells, respectively) from bone marrow or mature T cells, B cells, monocytes, and red blood cells (2×10^5, 1×10^5, 5×10^4, and 1×10^4 cells, respectively) were obtained for total RNA extraction. The RNA from each fraction was reverse-transcribed with an oligo dT primer. The cDNAs were approximately normalized based on a pilot β-actin amplification reaction. Aliquots of 2 μL of the direct or diluted cDNA pool were used in PCR of 18, 21, 24, 27, or 30 cycles with *MIXL*-EXP primers or β-actin primers as detailed in "Materials and methods." The amplification products (341 bp for *MIXL* and 246 bp for β-actin) were resolved on a 1.5% agarose gel, transferred to Hybond N+ membrane, and probed with radiolabeled *MIXL* or β-actin cDNAs. The blots were exposed to the Phosphor Screen for 1 hour. The images were converted from the Phosphor Screen. The number of PCR cycles is denoted at the bottom.

MIXL expression is restricted to progenitor compartments

The *MIXL* transcript levels in hematopoietic tissues were evaluated by Northern blotting analysis. We detected 3 transcript forms: 2.85 kb, 2.4 kb, and 1.9 kb (Figure 4A). It is likely that there are unidentified 5′ exons or different polyadenylation signals accounting for the 3 splice forms, as we could account for 1.7 kbp from the *MIXL* sequences and 3′-RACE results. All the 3 transcript forms were detected in lymph nodes, spleen, and bone marrow, suggesting that the *MIXL* expression is restricted to tissues with stem cell or progenitor compartments (Figure 4A). However, fetal liver did not reveal readily detectable expression, raising the possibility of a stage-specific low transcription during the 18- to 24-week stages of gestation. Future in situ hybridization experiments will address this important question.

To further assess *MIXL* expression during hematopoietic differentiation, we generated cDNA pools from fluorescence-activated cell sorter (FACS)–enriched progenitor populations of T, B, and myeloid lineages as well as mature B cells, T cells, monocytes and erythrocytes from normal peripheral blood. RT-PCR products of cDNA pools from these highly enriched fractions were detected by Southern hybridization with a radiolabeled *MIXL* probe. The products were screened at 18, 21, 24, 27, and 30 cycles to estimate the relative abundance of the target mRNA. The quality of the cDNA pools was first evaluated with a pair of β-actin primers. Under these conditions we could detect *MIXL* expression in both T- and B-cell progenitors from normal marrow (Figure 4B). In contrast, there were no detectable products from the RT-PCR of cDNA pools from mature T and B cells from peripheral blood (Figure 4B).

To further characterize MIXL protein expression, we developed polyclonal rabbit antisera against MIXL peptides. The specificities of the 2 antisera raised against 2 different epitopes were rigorously tested in transiently transfected 293T cells. Interestingly, the MIXL protein migrated at 36 kd, although the predicted molecular weight is 27 kd. Similarly, an expression construct lacking the homeodomain with a predicted molecular weight of 17 kd migrated at 26 kd (Figure 5A). The aberrant mobility could be due to the high proline content (35%) of the amino-terminal segment between amino acids 31 and 83. Both the transfected and endogenous MIXL proteins showed similar mobilities (Figure 5B) confirming the predicted ORF.

We next examined MIXL protein levels in cell lines derived from hematopoietic malignancies (Figure 5B). ML3 and KG1, AML cell lines that express $CD34^+$ antigen, expressed MIXL protein unlike the more committed myeloid leukemic cell lines HL60 and U937. The K562 cells derived from a patient with chronic myelogenous leukemia in erythroid blast crisis expressed significant levels. Among the cell lines derived from T-cell acute lymphocytic leukemia (ALL), the pre-T cell line CEM and the γ,δ^+ cell line Peer had higher levels than the more committed Jurkat cells. CJ, a lymphoma cell line derived from a patient with diffuse large B-cell lymphoma, expressed high levels of MIXL. Thus, the expression of *MIXL* appears to correlate with malignancies of early progenitors of myeloid, erythroid, and B- and T-cell lymphoid lineages.

MIXL induces of α-globin expression in *Xenopus* animal caps

The structural conservation between human *MIXL* and *Xenopus* Mix.1 led us to ask whether the human *MIXL* is involved in embryonic hematopoiesis. One way to address the question is to examine whether ectopic expression of *MIXL* would direct the

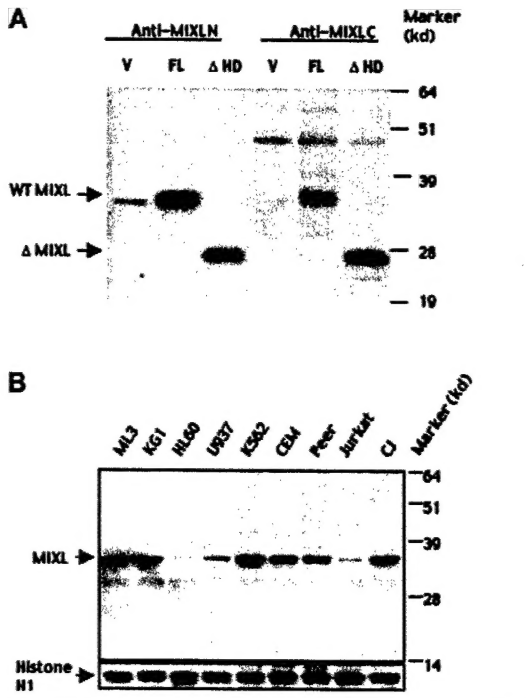

Figure 5. Characterization of MIXL protein expression. (A) Specificities of MIXL antibodies. 293T cells were transfected with wild-type (WT) *MIXL* and mutant *MIXL* without the homeobox (ΔMIXL) cloned in the vector pcDNA3.1-MycHis. Forty-eight hours after transfection, cytoplasmic lysates for homeodomain-less mutant MIXL and nuclear lysates for WT MIXL were resolved on a 12% SDS-PAGE gel and transferred to a Hybond P nylon membrane. The blot was probed with the affinity-purified antibody anti-*MIXL*-N at a dilution of 1:100 and antiserum anti-*MIXL*-C at a dilution of 1:1000. Both antibodies detected the same full-length MIXL protein at 36 kd and homeodomain-less mutant MIXL at 26 kd. V indicates vector alone; FL, full-length MIXL construct and ΔHD- homeodomain-less construct. (B) MIXL protein is expressed in hematologic malignancies. MIXL proteins were detected in leukemic cells expressing early myeloid markers (ML3, KG1), lymphoid markers (CEM, Peer, CJ) and erythroid markers (K562). The affinity purified antibody anti-*MIXL*-N was used at 1:100 dilution. The blot was reprobed with the mouse monoclonal antibodies against human histone H1 to quantify loading.

mesoderm to a hematopoietic fate in *Xenopus* embryos. Synthetic mRNA of wild-type *MIXL* or homeobox-truncated *MIXL* was injected into one-cell *Xenopus* embryos. At stage 8, animal caps were collected and cultured with human recombinant bFGF, because FGF or other mesoderm inducers are required for induction of hematopoietic mesoderm by BMP-4 or GATA-1 in animal cap assay.[13] At the equivalent of stage 36, animal caps were harvested for RNA extraction. As shown in Figure 6, bFGF alone could not induce αT4-globin expression in animal caps. With bFGF treatment, human *MIXL* induced α-globin expression in animal caps, whereas homeobox-truncated *MIXL* failed to induce erythropoiesis. These findings demonstrated that human *MIXL* mimics *Xenopus* Mix.1 in the specification and development of the hematopoietic system in *Xenopus* embryos. Taken together, *MIXL* may be a novel regulatory factor involved in both embryonic and adult human hematopoiesis.

Discussion

MIXL is a *Xenopus* Mix.1-like gene

The *Xenopus* Mix.1 homeoboxlike genes have been found in several organisms, including zebrafish, chicken, mouse, and human. Although at least 9 Mix-like genes have been isolated in

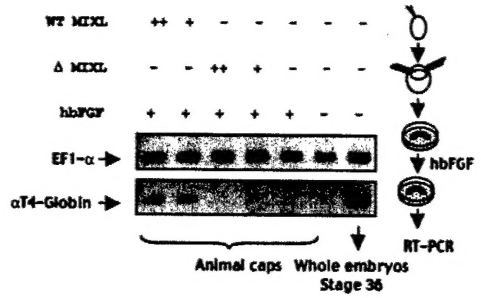

Figure 6. *MIXL*-induced expression of αT4-globin in *Xenopus* animal caps. Synthetic mRNAs (300 pg or 1 ng RNAs) were injected into *Xenopus* one-cell embryos in 2 separate sets. At the 8-cell stage, animal caps were excised and the explants were cultured in the presence of human recombinant bFGF (hbFGF). At the equivalent of stage 36, the animal caps from each set was pooled separately for RNA extraction and RT-PCR analysis. Twenty-seven cycles of multiplex PCR with the primer pairs for αT4-globin and EF1-α genes were performed. Note the αT4-globin expression in wild-type (WT) *MIXL*-injected animal caps but not in animal caps injected with homeobox-less *MIXL* (ΔMIXL) or FGF-treated embryos. +, 300 pg mRNA; ++, 1 ng mRNA.

Xenopus, only single-copy genes have been detected in each of the other vertebrates tested so far. How many Mix-like genes are there in higher organisms? Since *Xenopus* is tetraploid, there should be half the number of Mix-like genes in other organisms. However, our results strongly indicate that there is only one Mix-like gene in humans, because we could not identify any other Mix-like gene from the current genome database. Furthermore, Southern blotting studies with *MIXL* cDNA are consistent with *MIXL* being a single copy gene (H.L., W.G., and L.N., unpublished results, June, 1999).

A comparison between Mix-like genes indicated that the human *MIXL* gene is more closely related to the mouse *Mml* gene than other Mix-like genes. The homeodomain of human *MIXL* has 94% identity to that of the mouse *Mml*, whereas the human MIXL protein shows an overall 69% identity to mouse Mml. Although the overall homology between human *MIXL* and mouse *Mml* is not very high, it is likely that *MIXL* is the human ortholog of the mouse *Mml* because both genes localize to evolutionarily conserved syntenic regions on chromosome 1.[10]

Structural and functional conservation between Mix.1-like genes

Mix-like genes constitute a distinct family of pairedlike homeobox regulators. Typically, they encode a conserved Mix-like homeodomain and a conserved carboxy-terminal "WEE" motif, except for *Xenopus* Mixer that contains no "WEE" motif. However, there is a major difference in that Mix-like proteins in chicken, mouse, and human are much smaller in size (~230 amino acids) than those in amphibians.[8-11] *Xenopus* Mix-like genes encode proteins of approximately 370 amino acids in size.[3-7] There are 2 possible explanations that could account for the difference. The first is that only the conserved domains are evolutionarily critical for functions of Mix-like genes. Although homology between human *MIXL* and *Xenopus* Mix.1 is restricted in the conserved homeodomain and carboxy-terminal domain, ectopic expression of *MIXL* directed *Xenopus* mesoderm to a hematopoietic fate in animal cap assay (Figure 6), as *Xenopus* Mix.1 did. In addition, microinjection of *MIXL* mRNA into *Xenopus* embryos led to the phenotypic consequences similar to those of *Xenopus* Mix.1 (A.P.C. et al, unpublished results, March, 2000). This possibility can also explain why the overall homology between human *MIXL* and mouse *Mml* is relatively low compared to other pairedlike homeobox gene families. An alternate possibility is that *Mml*, *MIXL*, and *CMIX* may belong to a novel subfamily of Mix-like homeodomain proteins, in case there are additional Mix-like genes with weak homology in higher organisms.

MIXL in human hematopoiesis and leukemogenesis

Our results strongly suggest *MIXL* to be involved in human hematopoietic differentiation. Ectopic expression of *MIXL* could induce α-globin expression in animal caps (Figure 6), reflecting a possible role for human *MIXL* in embryonic hematopoiesis. Consistent with this observation was our detection of differential expression of *MIXL* in hematopoietic tissues and cell lineages (Figures 4 and 5). As shown in Figure 4A, the transcripts are readily detectable in lymph nodes and spleen. However, the apparent lack of expression in fetal liver is inconsistent with the expression in lymphoid tissues. The possibility of an oscillating expression pattern during hematopoietic maturation cannot be excluded. Clearly, the transcript levels are higher in the highly enriched B- and T-lymphoid progenitors from bone marrow than in the peripheral blood (Figure 4B). The high G+C content and the highly conserved homeodomain of *MIXL* are technical challenges in evaluating the expression pattern precisely. Future studies on how and when *MIXL* is induced in hematopoietic cells and whether *MIXL* cooperates with other pairedlike homeobox genes and transcription factors to regulate downstream events in hematopoiesis will elucidate the significance of its highly restricted expression.

Another indication of the highly restricted expression comes from the representation of *MIXL* gene in the human EST database. Among the 3.9 million entries from more than 30 tissue types and 20 cell lines *MIXL* is found in only 4 tissues, namely $CD20^+$ germinal center B cells, germ cell tumors, infiltrating ductal carcinoma, and fibrosarcoma.

BMP-4 signaling is hypothesized to be involved in initiation and specification of the mammalian hematopoietic system.[26,27] Although no direct evidence shows BMP-4 involvement in the mammalian hematopoietic system, *Xenopus* studies have implicated BMP-4 in hematopoiesis.[13,14,28] The hypothesis is further supported by the polarized expression of BMP-4 in a region of densely packed cells underlying intra-aortic hematopoietic clusters located in the AGM region.[15] Because *Xenopus* Mix.1 is a downstream target of the BMP-4 signaling pathway,[12] it will be interesting to determine if the human gene *MIXL* is a direct target of BMP-4.

Mammalian hematopoiesis requires many factors, including *SCL*, *LMO2*, and *GATA2*, that regulate the differentiation of uncommitted progenitors with high precision.[29-34] *Xenopus* SCL as well as GATA2 could induce the αT4-globin expression in animal caps.[35,36] With bFGF treatment, human *MIXL* induced αglobin expression in animal caps (Figure 6). Interestingly, *Xenopus* Mix.1 induces expression of *Xenopus* SCL,[36] suggesting that *MIXL* might be involved in *SCL* regulation. Considering the patterning of amphibian animal caps to hematopoietic fate by human *MIXL* (Figure 6) as well as *Xenopus* Mix.1,[12] future studies will unveil whether *MIXL* is involved in the same cascade in mammalian hematopoiesis by regulating *SCL*. Reagents generated in the present study will be of value in elucidating this mechanism.

Activation of early transcription factors, including homeobox genes, by chromosomal translocation is a recurrent theme in human hematologic malignancies (for a review, see van Oostveen et al[37]). We detect high levels of *MIXL* expression in some less-committed myeloid and lymphoid malignancies (Figure 5B). One possible mechanism for these observations is that high levels of *MIXL* expression may lead to activation and even overexpression of the

oncogene *SCL*. Additionally, aberrant expression of *MIXL* could directly confer a survival advantage or facilitate a block in differentiation.

Duplication or "jumping translocation" of 1q, resulting in increased gene dosage is considered a frequent secondary aberration in multiple myeloma.[25] Gain of chromosome 1q42 is also a common anomaly in breast cancer.[23-25] Thus, *MIXL* on chromosome 1q42 could be a novel candidate oncogene implicated in a subset of these anomalies.

Acknowledgments

We thank Dr Richard Ford for the CJ cell line, Elva Lopez for assistance with preparation of the manuscript, and members of our laboratory for critical reading of the manuscript. L.D.E. and L.N. are members of the Program in Genes and Development, Graduate School in Biomedical Sciences, University of Texas Health Science Center, Houston, TX.

References

1. Lebestky T, Chang T, Hartenstein V, Banerjee U. Specification of *Drosophila* hematopoietic lineage by conserved transcription factors. Science. 2000;288:146-149.
2. Lemaire P, Darras S, Caillol D, Kodjabachian L. A role for the vegetally expressed *Xenopus* gene Mix.1 in endoderm formation and in the restriction of mesoderm to the marginal zone. Development. 1998;125:2371-2380.
3. Rosa FM. Mix.1, a homeobox mRNA inducible by mesoderm inducers, is expressed mostly in the presumptive endodermal cells of *Xenopus* embryos. Cell. 1989;57:965-974.
4. Ecochard V, Cayrol C, Rey S , et al. A novel *Xenopus* mix-like gene milk involved in the control of the endomesodermal fates. Development. 1998; 125:2577-2585.
5. Henry GL, Melton DA. Mixer, a homeobox gene required for endoderm development. Science. 1998;281:91-96.
6. Mead PE, Zhou Y, Lustig KD, Huber TL, Kirschner MW, Zon LI. Cloning of Mix-related homeodomain proteins using fast retrieval of gel shift activities, (FROGS), a technique for the isolation of DNA-binding proteins. Proc Natl Acad Sci U S A. 1998;95:11251-11256.
7. Tada M, Casey ES, Fairclough L, Smith JC. Bix1, a direct target of *Xenopus* T-box genes, causes formation of ventral mesoderm and endoderm. Development. 1998;125:3997-4006.
8. Peale FV, Jr., Sugden L, Bothwell M. Characterization of CMIX, a chicken homeobox gene related to the *Xenopus* gene mix.1. Mech Dev. 1998;75:167-170.
9. Pearce JJ, Evans MJ. Mml, a mouse Mix-like gene expressed in the primitive streak. Mech Dev. 1999;87:189-192.
10. Robb L, Hartley L, Begley CG , et al. Cloning, expression analysis, and chromosomal localization of murine and human homologues of a *Xenopus* mix gene. Dev Dyn. 2000;219:497-504.
11. Stein S, Roeser T, Kessel M. CMIX, a paired-type homeobox gene expressed before and during formation of the avian primitive streak. Mech Dev. 1998;75:163-165.
12. Mead PE, Brivanlou IH, Kelley CM, Zon LI. BMP-4-responsive regulation of dorsal-ventral patterning by the homeobox protein Mix.1. Nature. 1996; 382:357-360.
13. Huber TL, Zhou Y, Mead PE, Zon LI. Cooperative effects of growth factors involved in the induction of hematopoietic mesoderm. Blood. 1998;92: 4128-4137.
14. Zhang C, Evans T. BMP-like signals are required after the midblastula transition for blood cell development. Dev Genet. 1996;18:267-278.
15. Marshall CJ, Kinnon C, Thrasher AJ. Polarized expression of bone morphogenetic protein-4 in the human aorta-gonad-mesonephros region. Blood. 2000;96:1591-1593.
16. Hejlik DP, Kottickal LV, Liang H , et al. Localization of SMAD5 and its evaluation as a candidate myeloid tumor suppressor. Cancer Res. 1997;57: 3779-3783.
17. Jiang Y, Liang H, Guo W, Kottickal LV, Nagarajan L. Differential expression of a novel C-terminally truncated splice form of SMAD5 in hematopoietic stem cells and leukemia. Blood. 2000;95:3945-3950.
18. Frohman MA. Rapid amplification of complementary DNA ends for generation of full- length complementary DNAs: thermal RACE. Methods Enzymol. 1993;218:340-356.
19. Cox DR, Burmeister M, Price ER, Kim S, Myers RM. Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. Science. 1990;250:245-250.
20. Harlow E, Lane D: Antibodies, A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory; 1988.
21. Kozak M. Initiation of translation in prokaryotes and eukaryotes. Gene. 1999;234:187-208.
22. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22:4673-4680.
23. Bieche I, Champeme MH, Lidereau R. Loss and gain of distinct regions of chromosome 1q in primary breast cancer. Clin Cancer Res. 1995;1: 123-127.
24. Buerger H, Otterbach F, Simon R, et al. Comparative genomic hybridization of ductal carcinoma in situ of the breast—evidence of multiple genetic pathways. J Pathol. 1999;187:396-402.
25. Sawyer JR, Tricot G, Mattox S, Jagannath S, Barlogie B. Jumping translocations of chromosome 1q in multiple myeloma: evidence for a mechanism involving decondensation of pericentromeric heterochromatin. Blood. 1998;91:1732-1741.
26. Li F, Lu S, Vida L, Thomson JA, Honig GR. Bone morphogenetic protein 4 induces efficient hematopoietic differentiation of rhesus monkey embryonic stem cells in vitro. Blood. 2001;98:335-342.
27. Schuldiner M, Yanuka O, Itskovitz-Eldor J, Melton DA, Benvenisty N. From the cover: effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. Proc Natl Acad Sci U S A. 2000;97:11307-11312.
28. Jones CM, Dale L, Hogan BL, Wright CV, Smith JC. Bone morphogenetic protein-4 (BMP-4) acts during gastrula stages to cause ventralization of *Xenopus* embryos. Development. 1996;122: 1545-1554.
29. Robb L, Lyons I, Li R , et al. Absence of yolk sac hematopoiesis from mice with a targeted disruption of the *scl* gene. Proc Natl Acad Sci U S A. 1995;92:7075-7079.
30. Shivdasani RA, Mayer EL, Orkin SH. Absence of blood formation in mice lacking the T-cell leukaemia oncoprotein tal-1/SCL. Nature. 1995;373: 432-434.
31. Tsai FY, Keller G, Kuo FC , et al. An early haematopoietic defect in mice lacking the transcription factor GATA-2. Nature. 1994;371:221-226.
32. Warren AJ, Colledge WH, Carlton MB, Evans MJ, Smith AJ, Rabbitts TH. The oncogenic cysteine-rich LIM domain protein rbtn2 is essential for erythroid development. Cell. 1994;78:45-57.
33. Yamada Y, Warren AJ, Dobson C, Forster A, Pannell R, Rabbitts TH. The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. Proc Natl Acad Sci U S A. 1998;95: 3890-3895.
34. Zhou Y, Lim KC, Onodera K , et al. Rescue of the embryonic lethal hematopoietic defect reveals a critical role for GATA-2 in urogenital development. EMBO J. 1998;17:6689-6700.
35. Maeno M, Mead PE, Kelley C , et al. The role of BMP-4 and GATA-2 in the induction and differentiation of hematopoietic mesoderm in *Xenopus laevis*. Blood. 1996;88:1965-1972.
36. Mead PE, Kelley CM, Hahn PS, Piedad O, Zon LI. SCL specifies hematopoietic mesoderm in *Xenopus* embryos. Development. 1998;125:2611-2620.
37. van Oostveen J, Bijl J, Raaphorst F, Walboomers J, Meijer C. The role of homeobox genes in normal hematopoiesis and hematological malignancies. Leukemia. 1999;13:1675-1690.